

Robust Regression Modelling using Generalized Maximum Likelihood Estimators and Generalized Scale Estimators on Indonesia's Poverty Depth Index

Yuliana Susanti^{1*}, Nuha Nurfauliyah², Kiki Ferawati³, Hasih Pratiwi⁴, Sri Sulistijowati Handajani⁵, Respatiwan⁶, Muhammad Bayu Nirwana⁷, and Andreas Rony Wijaya⁸

^{1,2,3,4,5,6,7,8}Statistics Department, Sebelas Maret University, Surakarta, Indonesia

*Corresponding author: yulianasusanti@staff.uns.ac.id

ABSTRACT

Poverty remains a critical issue that hinders both social and economic development by increasing the cost of growth and reducing the quality of human resources. One of the key indicators used to measure poverty is the Poverty Depth Index (PDI), which reflects the gap between poor individuals' expenditures and the poverty line. This study aims to model the Poverty Depth Index Indonesia in 2023 using robust regression with the Generalized Maximum Likelihood (GM) estimator and the Generalized Scale (GS) estimator. The model is designed to identify the factors influencing the PDI, including the percentage of the poor population, gini ratio, poverty line, and Human Development Index. Robust regression was applied because the presence of outliers in the data violated the normality assumption of classical linear regression. The results show that the GS estimator provides a better model with an AIC value of -39.5507 and an adjusted R-squared of 98.048% , compared to the GM estimator, which yielded an AIC of -10.8284 and an adjusted R-squared of 95.058% .

Keywords: Poverty Depth Index, Robust Regression, GM-estimation, GS-estimation.

1. INTRODUCTION

Indonesia is ranked fourth in the list of countries with the highest percentage of poor people in 2024. In the report, Indonesia is recorded as having 60.3 percent of its population living below the poverty line (Pratama, 2025). Poverty is one of the major issues that must be addressed and alleviated, as it affects various interconnected aspects of life. Poverty not only impacts social conditions but also hinders the economic development of a country by increasing the cost of economic growth. The effects on the economy can influence productivity levels, income, and the overall quality of human resources (Setiawan et al., 2024). Furthermore, the 2020–2024 National Medium-Term Development Plan identifies poverty alleviation as one of the priority agendas in national development efforts.

One of the key indicators used to measure the dimension of poverty is the Poverty Depth Index (PDI). The PDI is used to assess the gap in expenditure between poor individuals and the poverty line. Badan Pusat Statistik (BPS) states that an increase in the percentage of the poor population correlates with a rise in the PDI, as it indicates that more people are living further below the poverty line, thereby widening the disparity in expenditures among the poor (Badan Pusat Statistik, 2023). The poverty line serves as a threshold for determining the depth and severity of poverty. Changes in this threshold affect both the distribution and magnitude of the PDI. Research by Suryahadi and Sumarto (2003) shows that regions with high poverty depth

indices also tend to have high severity indices, indicating a more unequal distribution of expenditure among poor groups. A lack of access increases the risk of extreme poverty and worsens the conditions of the poor. Supporting this, a study by Agustin dkk. (2019) found that the Human Development Index (HDI) negatively affects the Poverty Depth Index. This indicates that improvements in human development significantly contribute to reducing the depth of poverty in the region.

The availability of accurate poverty data is essential in the strategy for poverty reduction, as it serves as a valid basis for government decision-making. This enables efforts to improve the potential of poor communities to be carried out more effectively and efficiently (Ferezagia, 2018). Regression analysis allows for the quantitative estimation of the effects of changes in independent variables (Gujarati & Porter, 2009). One commonly used method for parameter estimation in regression analysis is the Ordinary Least Squares (OLS). Classical regression models are based on several key assumptions, including normality, no autocorrelation, homoscedasticity, and no multicollinearity.

However, real-world data often exhibit characteristics such as the presence of outliers observations with values significantly different from the majority of the data. These outliers can greatly affect the validity of regression results (Dash et al., 2023). While outliers can distort the model if ignored, they also carry important information and should be properly addressed.

Robust regression is a statistical approach designed to provide reliable parameter estimates even in the presence of outliers or influential observations (Damayanti & Susanti, 2024). Among various robust methods, Generalized M-estimator (GM) and Generalized S-estimator (GS) offer enhanced resistance to outliers by incorporating both residual and leverage-based weighting mechanisms. These estimators improve the stability and accuracy of regression models under non-ideal conditions, where classical assumptions may not hold.

2. LITERATURE REVIEW

2.1 Poverty Depth Index

The Poverty Gap Index (P1) is a measure of the average shortfall in expenditure of poor individuals relative to the poverty line. A higher index value indicates that, on average, the poor are spending further below the poverty line. Poverty Depth Index in Indonesia fluctuated over the period from 2004 to 2023 and showed a similar pattern in both urban and rural areas. From 2004 to 2005, the index continuously declined but rose sharply in 2006. It then began to decrease again until 2014. Although it experienced increases in 2015, 2020, and 2021, the index declined once more in 2022 and 2023, in line with Indonesia's recovery from the Covid-19 pandemic (BPS, 2023).

2.2 Multiple Linear Regression

Linear regression analysis is a method that provides a model of the relationship between dependent variables and control variables $x_1, x_2, x_3, \dots, x_k$ with $\varepsilon_i \sim N(0, \sigma^2)$ (Montgomery et al., 2012). In general, the linear regression model for k independent variables can be written in Equation (1) as follows.

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i, \\ &= \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i, \quad i = 1, 2, 3, \dots, n; \quad j = 1, 2, \dots, k. \end{aligned} \quad (1)$$

with

- β_j : regression coefficient parameters to- j ,
- y_i : dependent variable in observation to- i ,
- x_i : independent variable in observation to- i ,
- ε_i : error value in data to- i ,
- n : amount of observation data,
- k : many independent variables.

If Equation (1) is written in matrix notation it will become,

$$y = X\beta + \varepsilon.$$

2.3 Normality Test

The normality test is used to test whether the data is normally distributed or not. According to Riag, et al. (2014), this test is a test of the normality of the error used to see whether the independent variables and dependent variables are normally distributed. The statistical test that is often used is the Kolmogorov-Smirnov Test (Gujarati, 2003).

1) Hypothesis Test

H_0 : Fallacy of normally distributed regression model

H_1 : The error of the regression model is not normally distributed

2) Level of Significance $\alpha = 0.05$

3) Critical Areas

H_0 rejected if value $D > D_{table}$ or $p - value < \alpha = 0.05$

4) Test Statistics

$$D = \max |F_0(X_i) - S_n(X_i)|$$

with:

$F_n(X)$: Normal Cumulative Probability

$F_0(X)$: Empirical Cumulative Probability

$S_n(X_i)$: Cumulative Frequency Distribution

5) Conclusion

If value $D > D_{table}$ or $p - value < \alpha = 0.05$ so H_0 rejected which means that the error is not normally distributed.

2.4 Outlier Identification

An outlier is an extreme value of an observation. Outliers affect the data analysis process. One measure to indicate whether an outlier is present or not is by calculating the Difference in Fitted Value (DFFITS) value. DFFITS is a measure of the influence caused by the to-i observation on the estimated value of $\hat{y}_i$ (Tiara, 2021). The DFFITS formula is as follows

$$(DFFITS) = t_i \sqrt{\left(\frac{h_{ii}}{1 - h_{ii}}\right)}, \quad i = 1, 2, \dots, n \quad (2)$$

with t_i is the studentized deleted residual for the to-i case and h_{ii} is the leverage value for the to-i data which is defined as $h_{ii} = X_i^t (X^t X)^{-1} X_i$ with $X_i^t = [1 \ x_{i1} \ x_{i2} \ \dots \ x_{ik}]$. A data is said to be an outlier if the value of $|(DFFITS)_i| > 2\sqrt{(k + 1)/n}$ with k is the number of independent variables.

2.5 Robust Regression

Robust regression is able to provide more stable estimates by reducing the influence of extreme observation data when the least squares method is used. The initial study on robust regression came from the Princeton Robustness Study conducted by Andrews et al. in 1972. The main purpose of robust regression is to provide stability of results against outliers. To achieve this stability, robust regression limits the influence of outliers.

2.6 Generalized Maximum Likelihood Estimation (GM)

Generalized maximum likelihood (GM) estimation is a development of maximum likelihood estimation. The basic idea behind GM estimation is to limit the influence of outliers on the variables x_i with weights w_i that just depends on x_i . According to Wilcox (2005), GM estimation is generally defined as,

$$\hat{\beta}_{GM} = \min \sum_{i=1}^n W_{iGM} \rho\left(\frac{e_i}{v(x_i)}\right), \quad (3)$$

with $v(x_i)$ is a weighting function for variables x_i . Estimator $\hat{\beta}$ the value obtained is not an invariant scale, so the value used is $\frac{e_i}{\hat{\sigma}}$ as a replacement e_i with $\hat{\sigma}$ is a scale factor that also needs to be estimated. Thus Equation (3) becomes,

$$\hat{\beta}_{GM} = \min \sum_{i=1}^n W_{iGM} \rho\left(\frac{y_i - \beta_0 - \sum_{j=1}^k x_{ij}\beta_j}{v(x_i)\hat{\sigma}_{GM}}\right). \quad (4)$$

Solve Equation (4) by lowering it against β and equated to zero, so that we get,

$$\sum_{i=1}^n W_{iGM} \psi\left(\frac{y_i - \beta_0 - \sum_{j=1}^k x_{ij}\beta_j}{v(x_i)\hat{\sigma}_{GM}}\right) x_{ij} = 0. \quad (5)$$

The weights used are Schweppe weights with $W_{iGM} = \sqrt{1 - h_{ii}}$ and $v(x_i) = W_{iGM}$ obtained,

$$W_{i,IGM} = W_{iGM} \frac{\psi\left(\frac{u_i}{v(X_i)}\right)}{(u_i)} = \frac{W_{iGM}}{u_i} \psi\left(\frac{u_i}{v(X_i)}\right) = \frac{\sqrt{1-h_{ii}}}{u_i} \psi\left(\frac{u_i}{\sqrt{1-h_{ii}}}\right), \quad (6)$$

with $u_{iGM} = \frac{e_i}{\hat{\sigma}_{GM}}$, $\hat{\sigma}_{GM} = 1,48M_l$, and M_l is the median of $(n-k)$ biggest of e_i . $\psi(x_i) = \max\{-K_{GM}, \min(K_{GM}, x)\}$ is the Huber influence function with $K_{GM} = 2\sqrt{\frac{(k+1)}{n}}$. Thus Equation (5) will become

$$\sum_i^n x_{ij} w_{iGM} (y_i - \beta_0 - \sum_{j=1}^k x_{ij} \beta_j) = 0, j = 1, 2, \dots, k. \quad (7)$$

The initial residuals used in GM estimation are the residuals obtained from OLS. Equation (7) can be solved using Iteratively Reweighted Least Squares (IRLS) until the value is obtained $\hat{\beta}$ which is convergent. The following is the GM estimation algorithm.

- Estimating regression model parameters with OLS.
- Calculating residual value $e_i = y_i - \hat{y}_i$ in the regression model with OLS.
- Calculating the median value of $(n-k)$ biggest of $|e_i| = M_l$.
- Calculate the estimated value $\hat{\sigma}_{GM} = 1,48M_l$.
- Calculating values $u_{iGM} = \frac{e_i}{\hat{\sigma}_{GM}}$.
- Calculating the weighting $W_{i,IGM}$ with function:

$$W_{i,IGM} = \frac{\sqrt{1-h_{ii}}}{u_i} \psi\left(\frac{u_i}{\sqrt{1-h_{ii}}}\right).$$

- Calculate parameter estimates $\hat{\beta}_{GM}$ with the weighted least squares method with $W_{i,IGM}$.
- Repeat steps (b) to (f) until a convergent $\hat{\beta}_{GM}$ value is obtained.

2.7 Generalized Scales Estimation (GS)

Quoting Saputera (2016) GS estimation is introduced as an extension of maximum likelihood estimation (M-estimation) to improve the robustness of scale estimation. It addresses the sensitivity of M-estimators to scale by utilizing a more resistant scale derive from pairwise residual differences. GS estimation can be defined following Croux et al. (1994), where the residuals are evaluated from a robust initial estimator, typically an M-estimator that has converged. For example $\hat{\beta}$ is an estimator β and $e(\hat{\beta}) = (e_1(\hat{\beta}), \dots, e_n(\hat{\beta}))$ is the residual vector with

$$e_i(\beta) = y_i - \sum_{j=0}^k x_{ij} \beta_j, 1 \leq i \leq n,$$

and $\Delta e_i(\beta) = (\Delta e_{12}(\beta), \dots, e_{n-1}(\beta))'$ is a vector of pairwise residual differences with $\Delta e_{ii}'(\beta) = e_i(\beta), \dots, e_i'(\beta)$ with $1 \leq i \leq i' \leq n$.

GS estimation can be defined as $\hat{\beta}_{GS}$ which has the following functions

$$\hat{\beta}_{GS} = \operatorname{argmin}_{\beta \in R^p} \hat{\sigma}_{GS}(\Delta e(\beta)) \quad (8)$$

where $\hat{\sigma}_{GS}$ obtained from the estimation of M differences of paired scale residuals which are solutions to the equation

$$\frac{1}{(n-2)} \sum_{i=1}^n \rho \left(\frac{\Delta e'_{ii}(\beta)}{\hat{\sigma}_{GS}} \right) = \delta \quad (\text{Croux et al. 1994}) \quad (9)$$

with ρ is the Tukey's Bisquare breaking function and δ is the breakdown point value. GS estimation $\hat{\beta}$ can be expressed in another form by decreasing to β , obtained as follows

$$\begin{aligned} \hat{\beta}_{GS} &= \operatorname{argmin}_{\beta \in R} p\sigma\hat{GS} \\ &= \frac{\sum_{i=1}^n \rho \left(\frac{\Delta e'_{ii}(\beta)}{\hat{\sigma}_{GS}} \right)}{\sum_{i=1}^n \rho' \left(\frac{\Delta e'_{ii}(\beta)}{\hat{\sigma}_{GS}} \right)} = \quad (10) \\ \sum_{i=1}^n \psi \left(\frac{\Delta e'_{ii}(\beta)}{\hat{\sigma}_{GS}} \right) &= 0 \end{aligned}$$

With ψ called the influence function which is a derivative of ρ . This form can be written the same as the S estimator with $w_i(u_i) = \frac{\psi(u_i)}{u_i}$ is an IRLS weighting function where $u_{iGS} = \frac{\Delta e'_{ii}}{\hat{\sigma}_{GS}}$ dan $C = 0,9958$. The initial residuals used in GS estimation are the pairwise residuals obtained from the least squares method. The process is continued with an iterative weighted OLS called Iterating Reweighted Least Square (IRLS) until convergence. GS estimation has the following algorithm:

- 1) Testing the classical assumptions of the OLS model.
- 2) Detecting outliers in the data.
- 3) Estimating the regression coefficients using M estimation.
 - a. Calculate the residuals of the regression model $e_i = y_i - \hat{y}_i$.
 - b. Calculate the scale estimate $\hat{\sigma}_M$ using the formula:

$$\hat{\sigma}_M = \frac{\operatorname{median} |e_i - \operatorname{median}(e_i)|}{0,6745}.$$

- c. Calculate u_{iM} using the formula:

$$u_{iM} = \frac{e_i}{\hat{\sigma}_M}.$$

- d. Calculate Tukey Bisquare weight w_{iM} using the function:

$$w_{iM}(u_{iM}) = \frac{\psi(u_{iM})}{u_{iM}} = \left\{ \left[1 - \frac{u_{iM}^2}{c_M^2} \right]^2, \quad |u_{iM}| < c_M, 0, \quad |u_{iM}| \geq c_M. \right.$$

With $c_M = 1,345$.

- e. Estimate the regression parameters $\hat{\beta}_M$ using the weights w_{iM} and iterate until convergence.
- 4) Calculate the pairwise residuals value of the M estimation model that has converged.

$$\Delta e_i(\beta) = \Delta e_{i12}(\beta), \dots, \Delta e_{(n-1)n}(\beta)$$

- 5) Calculate values $\hat{\sigma}_{GS}$

$$\hat{\sigma}_{GS} = \frac{\text{median } |e_i - \text{median}(e_i)|}{0,6745}.$$

6) Calculating values $u_{iGS} = \frac{\Delta e'_{ii}}{\hat{\sigma}_{GS}}$.

7) Calculating the weighted value w_i with the Tukey Bisquare weighting function

$$w(u_{iGS}) = \begin{cases} 1 - \left(\frac{u_{iGS}}{c}\right)^2 & |u_i| \leq c, 0 \\ & |u_i| > c, \end{cases}$$

with $c = 0.9958$.

8) Calculating the estimator $\hat{\beta}_S$ with w_i weighting.

9) Repeat the iteration until convergence is achieved, thus obtaining $\hat{\beta}_0^{GS}, \hat{\beta}_1^{GS}, \dots, \hat{\beta}_P^{GS}$ which is a GS estimate.

10) Calculating AIC dan R^2_{adj} values.

11) Determine the robust GS estimation model and interpret it.

3. METHOD

3.1 Data

The data used in this study comes from Badan Pusat Statistik Indonesia in 2023, consisting of Poverty Depth Index data for 34 provinces in Indonesia as the dependent variable (Y). The independent variables used are the percentage of poor people (X1), poverty line (X2), poverty severity index (X3), and the percentage of households with access to clean drinking water sources (X4).

3.2 Data Analysis

The research will be conducted using R-studio software with the analysis stages used in this study as follows.

- Estimating linear regression parameters using OLS.
- Testing the classical assumptions of the residuals of the human development index linear regression model with OLS.
- Identifying outliers if there is a violation of the normality assumption.
- Estimating robust regression parameters using GM estimation until convergence.
- Estimating robust regression parameters using GS estimation until convergence.
- Selecting the best model by comparing R^2_{adj} and AIC values.

4. RESULTS AND DISCUSSION

4.1 Parameter estimation with ordinary least square

In this study, the first step taken is to estimate the linear regression model using the least squares method (OLS). The following is the equation of the regression model with OLS.

$$\hat{y} = 1.678612 + 0.1981589x_1 + 4.999149x_2 + 0.0000019631x_3 - 0.00662588x_4. \quad (11)$$

Based on the research results using Rstudio software contained in Equation (11), R_{adj}^2 value is 0.96267, which means that 96.267% of the poverty depth index (y) in Indonesia in 2023 is influenced by the percentage of poor people (x_1), Gini ratio (x_2), poverty line (x_3), and human development index (x_4), while 3.733% is influenced by variables that are not included in the regression model. A valid regression model is determined through classical assumption tests that ensure the assumption criteria have been met. Based on the research results, the normality assumption tested with the Kolmogorov-Smirnov test resulted in a p-value of 0.0003893 which is less than $\alpha = 0,05$ which means that the residuals of the regression model are not normally distributed. If the assumption of normality is met, one of the steps that must be taken is to identify whether there are outliers or not. Outlier identification is done using the difference in fitted values (DFFITS). Data is said to be an outlier if the value of $|DFFITS| > 0.7669649888$. The identification results show that there are outliers in the 1st, 19th, 24th, 33rd, and 34th data.

Because the normality assumption was violated and there are five outliers, this study will use a robust regression model with GM estimation and GS estimation.

4.2 Generalized Maximum Likelihood Estimation

The use of robust regression GM estimation is carried out by limiting the influence of outliers on the variables. x_i with weights w_i which is done repeatedly to produce a value $\hat{\beta}_{GM}$ convergent.

Table 1. Results of Robust Regression Iteration GM Estimation

Iterati on	β_0	β_1	β_2	β_3	β_4
1	1.3511 11	0.1962 987	3.7207 08	0.0000001 5597	- 0.0526161
2	0.9836 21	0.1937 774	2.8748 20	0.0000001 1845	- 0.0403236
3	0.7663 01	0.1920 872	2.3574 49	0.0000000 9384	- 0.0327454
4	0.6404 79	0.1913 153	2.1922 26	0.0000000 8457	- 0.0294443
...	...	...	...	...	...
20	0.6013 54	0.1900 105	2.1521 07	0.0000000 7907	- 0.0281492
21	0.6013 54	0.1900 105	2.1521 08	0.0000000 7907	- 0.0281492
22	0.6013 54	0.1900 105	2.1521 08	0.0000000 7907	- 0.0281492

Based on Table 1, it is found that the parameters converge at the 22nd iteration. The following is the GM estimation robust regression model equation obtained for the Poverty Depth Index in Indonesia in 2023 as follows.

$$\hat{y} = 0.601354 + 0.190015x_1 + 2.152108x_2 + 0.0000007907x_3 + 0.0281492x_4. \quad (12)$$

Based on the results of the study using the Rstudio software contained in Equation (12), the R^2_{adj} value is 0.95768, which means that 95.768% of the poverty depth index (y) in Indonesia in 2023 is influenced by the percentage of poor people (x_1), the Gini ratio (x_2), the poverty line (x_3), and the human development index (x_4), while 4.232% is influenced by variables that are not included in the regression model.

4.3 Generalized Scales Estimation

Robust regression GS estimation begins by calculating the M estimates to produce $\hat{\beta}_M$ which converge in Table 2 as follows.

Table 2. Results of Robust Regression Iteration M Estimation

Iterati	β_0	β_1	β_2	β_3	β_4
on					
1	1.692205	0.1982430	4.977375	0.00000196561	-0.06672944
2	1.692455	0.1982551	4.976161	0.00000196580	-0.06672945
3	1.692450	0.1982560	4.976093	0.00000196580	-0.06672925
4	1.692449	0.1982561	4.976090	0.00000196580	-0.06672923
5	1.692449	0.1982561	4.976090	0.00000196580	-0.06672923

Table 2 shows that the values $\hat{\beta}_M$ which converges is obtained in the 5th iteration. Next, GS estimation is carried out starting by using the residue from the converged M estimate value. The results of the GS estimation iteration are presented in Table 3.

Table 3. Robust Regression Iteration Results of GS Estimation

Iterati	β_0	β_1	β_2	β_3	β_4
on					
1	1.482618	0.1993011	2.971987	0.0000012280	-0.00483829
2	1.216670	0.1989314	2.370755	0.0000010307	-0.00402736

3	0.9955 64	0.1989 458	1.9372 78	0.000000 8670	- 0.00339255
...	...	...	...	...	...
45	0.4032 962	0.1725 109	0.9893 16	0.000000 1884	- 0.01354644
46	0.4032 961	0.1725 109	0.9893 16	0.000000 1884	- 0.01354644
47	0.4032 961	0.1725 109	0.9893 16	0.000000 1884	- 0.01354644

Based on Table 3, it is found that the parameters converge at the 47th iteration. The following is the equation of the GS estimation robust regression model obtained for the Poverty Depth Index in Indonesia in 2023 as follows.

$$\hat{y} = 0.4032961 + 0.1725109x_1 + 0.989316x_2 + 0.0000001884x_3 - 0.013546 \quad (13)$$

Based on the results of the study using the Rstudio software contained in Equation (13), the R_{adj}^2 value is 0.98048, which means that 98.048% of the poverty depth index (y) in Indonesia in 2023 is influenced by the percentage of poor people (x_1), the Gini ratio (x_2), the poverty line (x_3), and the human development index (x_4), while 1.952% is influenced by variables that are not included in the regression model.

4.4 Best Selection Model

The best model is selected based on the smallest Akaike Information Criterion (AIC) and the biggest R_{adj}^2 value. The Adjusted determination coefficient values and AIC values for the robust regression of GM estimation and GS estimation are shown in Table 4.

Table 4. Comparison of Robust Regression GM Estimation and GS Estimation

Model	Adjusted Determination Coefficient	AIC
GM Estimation	95.768%	- 10.82844861
GS Estimation	98.048%	- 39.55074099

Table 4 shows that the GS estimation robust regression model has a larger Adjusted determination coefficient value and a smaller AIC value compared to the GM estimation robust regression model. Therefore, it can be concluded that the GS estimation robust regression model is the best model used to estimate the poverty depth index in 2023.

5. CONCLUSION

From the research results, it was found that the best model for Indonesia's poverty depth index in 2023 is as follows.

$$\hat{y} = 0.4032961 + 0.1725109x_1 + 0.989316x_2$$

$$+0.0000001884x_3-0.01354644 x_4,$$

with the interpretation that every increase in the percentage of poor people (x_1) will increase the poverty depth index by 0.1725109; every increase in the gini ratio (x_2) will increase the poverty depth index by 0.9899316; the poverty line (x_3) will increase the poverty depth index by 0.0000001884; and the human development index (x_4) will decrease the poverty depth index by 0.01354644. The model has an adjusted R-Square value of 0.98048, which means that 98.048% of the poverty depth index (y) in Indonesia in 2023 is influenced by the percentage of poor people (x_1), the gini ratio (x_2), the poverty line (x_3), and the human development index (x_4) while 1.952% is influenced by variables not included in the regression model.

6. ACKNOWLEDGEMENT

This research was supported by RKAT Universitas Sebelas Maret for the Fiscal Year 2025 through the Research Grant Scheme PENGUATAN KAPASITAS GRUP RISET (PKGR-UNS) Category B, under the Research Assignment Agreement Number: 371/UN27/PT.01.03/2025. The authors gratefully acknowledge this support, which has significantly contributed to the successful completion of our research.

The authors would also like to express our sincere gratitude to Universitas Sebelas Maret for providing the facilities and academic environment that supported this work. Special thanks go to our research group members for valuable insights, constructive feedback, and continuous encouragement throughout the research process.

REFERENCES

- Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., & Tukey, J. W. (1972). *Robust estimates of location: Survey and advances*. Princeton University Press.
- Agustin, N., Hidayat, M. S., & Umiyati, E. (2019). Analisis pengaruh pertumbuhan ekonomi, indeks pembangunan manusia (IPM) dan upah minimum provinsi (UMP) terhadap kemiskinan di Kabupaten Merangin. *E-Jurnal Perspektif Ekonomi dan Pembangunan Daerah*, 8(1), 53–64. <https://doi.org/10.22437/pdpd.v8i1.7346>.
- Badan Pusat Statistik. 2023. *Penghitungan dan Analisis Kemiskinan Makro Indonesia Tahun 2023*. Jakarta: Badan Pusat Statistik.
- Croux, C., Rousseeuw, P. J., & Hossjer, O. (1994). Generalized S-estimators. *Journal of the American Statistical Association*, 89(428), 1271–1281. <https://doi.org/10.2307/2290986>
- Damayanti, A. D., & Susanti, Y. (2024). Pemodelan regresi robust estimasi GM pada data sosial ekonomi. *Prosiding Seminar Nasional Statistika dan Data Sains*.

- Dash, S. S., Sahoo, S., & Sethi, N. (2023). Estimating the impact of HDI on economic growth: Evidence from South Asian countries. *Journal of Economic Studies*, 50(4), 701–716. <https://doi.org/10.1108/JES-06-2022-0312>
- Ferezagia, M. (2018). *Analisis pengaruh indeks pembangunan manusia terhadap pertumbuhan ekonomi di Indonesia*. Skripsi. Universitas Negeri Semarang.
- Gujarati, D. N. (2003). *Basic econometrics* (4th ed.). McGraw-Hill.
- Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics* (5th ed.). McGraw-Hill.
- International Labour Organization. (2005). *Chapter 4: Measures of Poverty*. Poverty Manual, All, JH Revision of August 8, 2005. Retrieved from https://www.ilo.org/sites/default/files/wcmsp5/groups/public/%40americas/%40ro-lima/%40sro-port_of_spain/documents/presentation/wcms_304851.pdf
- Montgomery, Do. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis*. John Wiley & Sons, Inc.
- Pratama, R. L.. (2025). *Indonesia peringkat 4 negara dengan penduduk miskin terbanyak versi Bank Dunia*. Kompas TV. Retrieved from <https://www.kompas.tv/ekonomi/592311/indonesia-peringkat-4-negara-dengan-penduduk-miskin-terbanyak-versi-bank-dunia>.
- Riang, M. F., Adhariani, D., & Susanti, Y. (2014). Analisis regresi robust GM untuk pendugaan parameter pada data ekonomi dengan pencilan. *Jurnal Statistika*, 10(1), 35–42.
- Saputera, Y. H. (2016). *Penerapan regresi robust Generalized S-Estimator pada data IPM Indonesia*. Skripsi Universitas Diponegoro.
- Setiawan, M., Dwiatmanto, & Permatasari, A. D. (2024). Analisis regresi robust pada data indeks pembangunan manusia. *Jurnal Ekonomi dan Pembangunan Daerah*, 12(2), 110–120.
- Suryahadi, A., & Sumarto, S. (2003). Poverty and vulnerability in Indonesia before and after the economic crisis. *Asian Economic Journal*, 17(1), 45–64.
- Tiara, R. A. (2021). *Analisis faktor-faktor yang memengaruhi indeks pembangunan manusia di Provinsi Jawa Barat*. Skripsi Universitas Pendidikan Indonesia.
- Wilcox, R. R. (2005). *Introduction to Robust Estimation and Hypothesis Testing* (2nd ed). San Diego: Academic Press.